Problem Set 6 Solution
by Yuan Zhou

1. Let $\boldsymbol{w}^*$ be the vector such that $\|\boldsymbol{w}^*\| = 1$ and $\forall i, (\boldsymbol{w}^* \cdot \boldsymbol{a}_i) l_i \geq \gamma$.

   As the algorithm runs, we keep track of the value $\boldsymbol{w} \cdot \boldsymbol{w}^*$ and $\|\boldsymbol{w}\|$.

   At each iteration, when we update the vector $\boldsymbol{w}$ by $\boldsymbol{w}' = \boldsymbol{w} + l_i \boldsymbol{a}_i$, we have

$$\boldsymbol{w}' \cdot \boldsymbol{w}^* = \boldsymbol{w} \cdot \boldsymbol{w}^* + l_i(\boldsymbol{a}_i \cdot \boldsymbol{w}^*) \geq \boldsymbol{w} \cdot \boldsymbol{w}^* + \gamma,$$

   and

$$\|\boldsymbol{w}'\| = \|\boldsymbol{w} + l_i \boldsymbol{a}_i\| = \sqrt{\|\boldsymbol{w}\|^2 + \|\boldsymbol{a}_i\|^2 + 2l_i(\boldsymbol{w} \cdot \boldsymbol{a}_i)} \leq \sqrt{\|\boldsymbol{w}\|^2 + 1 + \gamma\|\boldsymbol{w}\|}.$$

   We now show by induction that at step t, we have $\|\boldsymbol{w}^{(t)}\| \leq \frac{3}{\gamma} + \frac{2\gamma}{3} \cdot t$.

   - When $t = 0$, we have $\|\boldsymbol{w}^{(0)}\| = 0 \leq \frac{3}{\gamma}$.
   - For $t > 0$, we have

$$\begin{aligned}\|\boldsymbol{w}^{(t)}\| \leq& \sqrt{\|\boldsymbol{w}^{(t-1)}\|^2 + 1 + \gamma\|\boldsymbol{w}^{(t-1)}\|} \\ \leq& \sqrt{\left(\frac{3}{\gamma} + \frac{2\gamma}{3} \cdot (t-1)\right)^2 + 1 + \gamma\left(\frac{3}{\gamma} + \frac{2\gamma}{3} \cdot (t-1)\right)} \qquad (*).\end{aligned}$$

     Note that since

$$\begin{aligned}&\left(\frac{3}{\gamma} + \frac{2\gamma}{3} \cdot t\right)^2 - \left(\frac{3}{\gamma} + \frac{2\gamma}{3} \cdot (t-1)\right)^2 - 1 - \gamma\left(\frac{3}{\gamma} + \frac{2\gamma}{3} \cdot (t-1)\right) \\ =&\frac{2\gamma}{3}\left(\frac{6}{\gamma} + \frac{2\gamma}{3} \cdot (2t-1)\right) - 1 - \gamma\left(\frac{3}{\gamma} + \frac{2\gamma}{3} \cdot (t-1)\right) \\ =&\frac{4\gamma^2}{9} \cdot (2t-1) - \gamma \cdot \frac{2\gamma}{3} \cdot (t-1) \\ =&\left(\frac{8}{9} - \frac{2}{3}\right)\gamma^2 t + \left(\frac{2}{3} - \frac{4}{9}\right)\gamma^2 \geq 0,\end{aligned}$$

     we have

$$(*) \leq \sqrt{\left(\frac{3}{\gamma} + \frac{2\gamma}{3} \cdot t\right)^2} = \frac{3}{\gamma} + \frac{2\gamma}{3} \cdot t.$$

     This completes the induction.

   Therefore, at step t, we have

$$\frac{\boldsymbol{w} \cdot \boldsymbol{w}^*}{\|\boldsymbol{w}\|} \geq \frac{t \cdot \gamma}{\frac{3}{\gamma} + \frac{2\gamma}{3} \cdot t} = \frac{1}{\frac{3}{t \cdot \gamma^2} + \frac{2}{3}}.$$

   When $t > 9/\gamma^2$, this value is greater than 1. On the other hand, since $\frac{\boldsymbol{w} \cdot \boldsymbol{w}^*}{\|\boldsymbol{w}\|}$ is the cosine value of the angle between $\boldsymbol{w}$ and $\boldsymbol{w}^*$ and cannot be greater than 1, we know that the algorithm terminates within $9/\gamma^2$ steps.

2. 7. (Proof omitted).

3. Let $A \subseteq U$ be a set that is shattered such that $|A| = d$. Suppose that $A = \{a_1, a_2, \ldots, a_d\}$ and $\epsilon < 1/2$.

   Consider the following probability distribution

$$p(a_i) = \begin{cases} \frac{4\epsilon}{d} & \text{for } i \leq d/2 \\ 0 & \text{for } d/2 < i \leq d-1 \\ 1-2\epsilon & \text{for } i = d \end{cases} .$$

   Note that for every $S \subseteq \{a_1, a_2, \ldots, a_{d/2}\}$ such that $|S| \geq d/4$, we have that $p(S) \geq \epsilon$. Therefore, there exists $S' \subseteq \mathcal{F}$ such that $S' \cap A = S$ (therefore $p(S') = p(S) \geq \epsilon$ by the definition of p).

   Since we need to hit all the sets which intersect $\{a_1, a_2, \ldots, a_{d/2}\}$ with at least $d/4$ elements, we need to choose at least $d/4$ elements from $\{a_1, a_2, \ldots, a_{d/2}\}$. It is easy to see that $\Omega(d/\epsilon)$ samples are needed.

4. (a) n. (Proof omitted.)

   (b) Suppose g is the unknown conjunction function being learned. For any conjunction function f, if the algorithm sees a sample x such that $f(x) \neq g(x)$, f cannot be the final output of the algorithm. Therefore, if f is ϵ-far from g, at each sample, with probability ϵ, f is "killed". In total, if f is ϵ-far from g, f survives with probability at most $(1-\epsilon)^m$ where m is the number of samples the algorithm uses.

   Since there are at most 3^n possible conjunction functions (therefore this is also an upper bound for the number of conjunction functions that are ϵ-far from g), by a union bound, the probability at a function ϵ-far from g survives is at most $(1-\epsilon)^m \cdot 3^n$. By making $m = c \cdot \frac{1}{\epsilon}(n + \log(1/\delta)$ for some large enough c, we are able to make the probability at most δ.

5. (a) Since the dimension of v_F is $|\mathcal{S}_d|$, the number of independent vectors $|\{v_F : F \in \mathcal{F}\}| = |\mathcal{F}|$ is at most $|\mathcal{S}_d| = \sum_{i=0}^{d} \binom{n}{i}$.

   (b) Note that we have $\sum_{F \in \mathcal{F}} \alpha_F v_F = 0$. At coordinate X, we have $\sum_{F \in \mathcal{F}} \alpha_F v_F(X) = 0$. Since $v_F(X) = 1$ when $X \subseteq F$ and $v_F(X) = 0$ otherwise, we get

$$\mu(X) = \sum_{F \in \mathcal{F} : X \subseteq F} \alpha_F = 0.$$

   (c) Since α is not a 0 vector, let T be the maximum-sized set in $\mathcal{F}$ such that $\alpha_T \neq 0$. Observe that

$$\mu(T) = \sum_{F \in \mathcal{F} : T \subseteq F} \alpha_F = \alpha_T \neq 0.$$

(d) Observe that

$$\begin{aligned}
&\sum_{W:Z\subseteq W\subseteq Y}(-1)^{|W\setminus Z|}\mu(W)\\
&=\sum_{W:Z\subseteq W\subseteq Y}(-1)^{|W\setminus Z|}\sum_{F\in\mathcal{F}:W\subseteq F}\alpha_F\\
&=\sum_{F\in\mathcal{F}:Z\subseteq F}\alpha_F\sum_{W:Z\subseteq W\subseteq Y,W\subseteq F}(-1)^{|W\setminus Z|}\\
&=\sum_{F\in\mathcal{F}:Z\subseteq F}\alpha_F\sum_{W:Z\subseteq W\subseteq Y\cap F}(-1)^{|W\setminus Z|},
\end{aligned}$$

where for $Z\subseteq Y\cap F$, $\sum_{W:Z\subseteq W\subseteq Y\cap F}(-1)^{|W\setminus Z|}=1$ only when $Z=Y\cap F$ and it equals 0 otherwise. Therefore, we have

$$\sum_{W:Z\subseteq W\subseteq Y}(-1)^{|W\setminus Z|}\mu(W)=\sum_{F\in\mathcal{F}:Z\subseteq F}\alpha_F\sum_{W:Z\subseteq W\subseteq Y\cap F}(-1)^{|W\setminus Z|}=\sum_{F\in\mathcal{F}:Y\cap F=Z}\alpha_F.$$

(e) Since Y is the set with smallest cardinality such that $\mu(Y)\neq 0$, we have $\mu(W)=0$ for all $W\subsetneq Y$. Thus,

$$\sum_{W:Z\subseteq W\subseteq Y}(-1)^{|W\setminus Z|}\mu(W)=(-1)^{|Y\setminus Z|}\mu(Y)\neq 0.$$

Therefore, by part (d), we have

$$\sum_{F\in\mathcal{F}:Y\cap F=Z}\alpha_F\neq 0.$$

Therefore, there exists at least one $F\in\mathcal{F}$ such that $Y\cap F=Z$.

(f) By part (e), Y is shattered by $\mathcal{F}$. By part (b), we have that $|Y|>d$. Therefore, the VC dimension of $\mathcal{F}$ is at least $d+1$ – a contradiction.